

5.00 crédits

30.0 h + 15.0 h

Q2

Enseignants	Caelen Olivier ;Heuchenne Cédric ;
Langue d'enseignement	Anglais
Lieu du cours	Louvain-la-Neuve
Préalables	Connaissances en programmation (ex. LDATS2030)
Thèmes abordés	Ce cours couvre les composantes conceptuelles et pratiques essentielles de la science des données appliquée. Les principaux thèmes abordés sont les suivants : – Domaines d'application de la science des données et études de cas concrets illustrant le développement d'un projet de bout en bout. – Cycle de vie standard d'un projet de science des données : acquisition des données, analyse exploratoire, nettoyage et prétraitement des données, développement, évaluation et déploiement du modèle. – Les principes fondamentaux de l'apprentissage supervisé, avec un accent particulier sur les problèmes de classification et de régression. – Une introduction aux modèles classiques d'apprentissage automatique, notamment les arbres de décision, les k plus proches voisins et les réseaux neuronaux... – Une mise en œuvre pratique à l'aide de l'écosystème scientifique Python et d'outils modernes prenant en charge des pipelines et un déploiement reproductibles.
Acquis d'apprentissage	A la fin de cette unité d'enseignement, l'étudiant est capable de : En ce qui concerne le cadre des AA du Master [120] en science des données : statistiques, cette activité contribue au développement et à l'acquisition des AA suivants : • en priorité : 1.1 – 1.2 – 1.3 – 2.1 – 2.4 – 2.5 – 6.3 • en secondaire : 4.1 – 4.2 – 4.4 – 5.3 – 5.6 – 6.1 ¹ En ce qui concerne le cadre des AA du Master [120] en Statistiques : Général, cette activité contribue au développement et à l'acquisition des AA suivants : • en priorité : 1.3 – 2.2 – 2.5 – 3.3 – 5.3 • en secondaire : 2.4 – 5.4 – 5.6 – 6.3
Modes d'évaluation des acquis des étudiants	Projet de groupe et examen oral individuel. L'examen oral peut inclure des questions sur le projet. Les deux parties sont obligatoires pour réussir le cours.
Méthodes d'enseignement	Le cours combine des cours magistraux appuyés par des diapositives et des sessions pratiques sur ordinateur au cours desquelles les étudiants appliquent chaque concept à l'aide de notebooks Python.
Contenu	1. Introduction à la science des données 1. Aperçu du processus de la science des données : compréhension des activités, compréhension des données, modélisation, évaluation, déploiement. 2. Étude de cas 2. Extraction et manipulation de données en Python 1. Introduction à l'écosystème scientifique Python : NumPy, pandas, ... 2. Chargement de données à partir de fichiers et d'API (CSV, JSON, SQL) 3. Prétraitement des données : valeurs manquantes, ingénierie des caractéristiques, encodage des variables catégorielles, mise à l'échelle et normalisation ... 4. Analyse exploratoire et visualisation des données (matplotlib, seaborn,...) 3. Apprentissage supervisé : classification et régression 1. Introduction à l'écosystème d'apprentissage automatique Python : sklearn & statsmodels 2. Modèles classiques d'apprentissage automatique pour les données tabulaires : • K plus proches voisins • arbres de décision et forêt aléatoire • introduction aux réseaux neuronaux et PyTorch

	4. Évaluation et interprétabilité des modèles 1. Divisions train/validation/test, validation croisée, mesures de performance (précision, F1, ROC-AUC, MSE) 2. Importance des caractéristiques et explicabilité des modèles (SHAP, importance des permutations) 5. Introduction aux outils et à l'environnement informatique 1. Jupyter, MLflow, Streamlit, FastAPI
Bibliographie	"Hands#On Machine Learning with Scikit#Learn and PyTorch: Concepts, Tools, and Techniques to Build Intelligent Systems", Aurélien Géron, O'Reilly Media, ISBN -13: 979-8341607989 "Machine Learning with PyTorch and Scikit-Learn: Develop machine learning and deep learning models with Python", Sebastian Raschka, Yuxi (Hayden) Liu & Vahid Mirjalili, Packt Publishing, ISBN-13: 978-1801819312.
Faculté ou entité en charge:	LSBA

Programmes / formations proposant cette unité d'enseignement (UE)				
Intitulé du programme	Sigle	Crédits	Prérequis	Acquis d'apprentissage
Master [120] en science des données, orientation statistique	DATS2M	5		
Master [120] en statistique, orientation biostatistiques	BSTA2M	5		
Master [120] en linguistique	LING2M	5		
Master [120] : bioingénieur en sciences et technologies de l'environnement	BIRE2M	5		
Master [120] en sciences mathématiques	MATH2M	5		
Master [120] en sciences actuarielles	ACTU2M	5		
Master [120] en statistique, orientation générale	STAT2M	5		
Master [120] : bioingénieur en chimie et bioindustries	BIRC2M	5		
Master [120] : ingénieur civil en mathématiques appliquées	MAP2M	5		
Mineure en statistique, sciences actuarielles et science des données	MINSTAT	5		
Certificat d'université : Statistique et science des données (15/30 crédits)	STAT2FC	5		